Journal of Pathology
J Pathol 2021
Published online 24 November 2021 in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/path.5819

ORIGINAL PAPER

A digital score of tumour-associated stroma infiltrating lymphocytes predicts survival in head and neck squamous cell carcinoma

Muhammad Shaban[1], Shan E Ahmed Raza[1], Mariam Hassan[2], Arif Jamshed[2], Sajid Mushtaq[2], Asif Loya[2], Nikolaos Batis[3], Jill Brooks[3], Paul Nankivell[3], Neil Sharma[3], Max Robinson[4], Hisham Mehanna[3], Syed Ali Khurram[5] and Nasir Rajpoot[1,6,7]*

[1] *Department of Computer Science, University of Warwick, Coventry, UK*
[2] *Department of Pathology, Shaukat Khanum Memorial Cancer Hospital Research Centre, Lahore, Pakistan*
[3] *Institute of Head and Neck Studies and Education, University of Birmingham, Birmingham, UK*
[4] *School of Dental Sciences, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK*
[5] *School of Clinical Dentistry, University of Sheffield, Sheffield, UK*
[6] *The Alan Turing Institute, London, UK*
[7] *Department of Pathology, University Hospitals Coventry & Warwickshire NHS Trust, Coventry, UK*

**Correspondence to: N Rajpoot, Department of Computer Science, University of Warwick, Coventry, CV4 7AL, UK. E-mail: n.m.rajpoot@warwick.ac.uk*

Abstract

The infiltration of T-lymphocytes in the stroma and tumour is an indication of an effective immune response against the tumour, resulting in better survival. In this study, our aim was to explore the prognostic significance of tumour-associated stroma infiltrating lymphocytes (TASILs) in head and neck squamous cell carcinoma (HNSCC) through an AI-based automated method. A deep learning-based automated method was employed to segment tumour, tumour-associated stroma, and lymphocytes in digitally scanned whole slide images of HNSCC tissue slides. The spatial patterns of lymphocytes and tumour-associated stroma were digitally quantified to compute the tumour-associated stroma infiltrating lymphocytes score (TASIL-score). Finally, the prognostic significance of the TASIL-score for disease-specific and disease-free survival was investigated using the Cox proportional hazard analysis. Three different cohorts of haematoxylin and eosin (H&E)-stained tissue slides of HNSCC cases ($n = 537$ in total) were studied, including publicly available TCGA head and neck cancer cases. The TASIL-score carries prognostic significance ($p = 0.002$) for disease-specific survival of HNSCC patients. The TASIL-score also shows a better separation between low- and high-risk patients compared with the manual tumour-infiltrating lymphocytes (TILs) scoring by pathologists for both disease-specific and disease-free survival. A positive correlation of TASIL-score with molecular estimates of $CD8^+$ T cells was also found, which is in line with existing findings. To the best of our knowledge, this is the first study to automate the quantification of TASILs from routine H&E slides of head and neck cancer. Our TASIL-score-based findings are aligned with the clinical knowledge, with the added advantages of objectivity, reproducibility, and strong prognostic value. Although we validated our method on three different cohorts ($n = 537$ cases in total), a comprehensive evaluation on large multicentric cohorts is required before the proposed digital score can be adopted in clinical practice.

Keywords: digital pathology; deep learning; tumour-associated stroma; survival analysis; head and neck squamous cell carcinoma; artificial intelligence; machine learning

Received 28 March 2021; Revised 1 October 2021; Accepted 23 October 2021

No conflicts of interest were declared.

Introduction

There are around 900 000 annual new cases of head and neck cancers worldwide and 450 000 annual deaths [1]. Head and neck squamous cell carcinoma (HNSCC) accounts for approximately 90% of head and neck cancers and is the sixth leading cancer by incidence worldwide [2]. HNSCC predominantly develops in the epithelial lining of the oral cavity, sinonasal tract, pharynx, and larynx [3]. Major risk factors include tobacco smoking, tobacco chewing, alcohol consumption, and human papillomavirus infection [4,5]. The prognosis of HNSCC remains poor, with a 28–67% chance of survival at 5 years [6], highlighting the need for novel biomarkers and objective quantitative analysis of any potential prognostic markers to stratify patients

into appropriate risk groups and identify those who may benefit from aggressive treatment from those that can be put under surveillance [7].

Tumour-infiltrating lymphocytes (TILs) have been shown to be of prognostic significance for HNSCC [8,9]. TILs are not routinely quantified in diagnostic practice, although some methods for manual TIL scoring on haematoxylin and eosin (H&E)-stained tissue sections have been reported in the literature. However, this quantification process is subjective and prone to inter- and intra-observer variability. Recently, the International Immuno-Oncology Biomarker Working Group has developed guidelines for TIL assessment in breast cancer [10] to standardise and obtain a more reproducible and objective TIL score. However, there are no such guidelines for TIL assessment in HNSCC. Therefore, a computer-based automated method may help to eliminate the subjectivity through objective TIL quantification and assist with prognostic stratification.

The emerging area of computational pathology has seen a surge of interest in recent years, with a variety of algorithms proposed in the literature for detection of lymph node metastasis in breast cancer [11], tumour detection [12] and segmentation [13–15], cancer grading [16,17], and prognostics [18–20], to list only a few. There have been some studies on automated quantification of lymphocytic infiltration in whole slide images (WSIs) of H&E-stained histology tissue slides of different cancers. For instance, Saltz *et al* [21] quantified the spatial patterns of lymphocytes in WSIs, independent of tumour location, to investigate their prognostic significance in 14 different cancer types. Maley *et al* [22] reported co-localisation between immune cells and cancer cells as a prognostic factor for breast cancer. Nawaz *et al* [23] used hotspot analysis to identify the statistically significant regions of cancer and immune cells. Shaban *et al* [19] quantified the abundance of TILs for disease-free survival (DFS) analysis of oral squamous cell carcinoma (OSCC) patients. Most existing automated quantification methods either only consider lymphocytes or lymphocytic infiltration in tumour regions. However, some clinical studies [24,25] have reported the prognostic significance of lymphocyte infiltration in tumour-associated stroma (TAS) in HNSCC.

We propose a novel objective quantification method of lymphocytic infiltration in TAS (see Figure 1). The proposed method calculates the percentage of TAS co-localised with lymphocytes, which we term the TASIL-score. We evaluate the prognostic significance on three independent patient cohorts ($n = 537$ cases in total). The proposed TASIL-score shows prognostic significance ($p = 0.002$) for disease-specific survival (DSS) of HNSCC patients. The TASIL-score is also a prognostic indicator for DSS and DFS in two OSCC and oropharyngeal squamous cell carcinoma (OPSCC) cohorts. We have also compared the predictive ability of a TASIL-score-based survival model with existing quantification methods through a concordance index measure where the TASIL-score achieved the highest concordance compared with its counterparts.

Materials and methods

Clinical samples

Three different patient cohorts (TCGA-HN [26], SKM [19], and PredicTR1 [27]) were used in this study. The TCGA-HN cohort (C1) is a publicly available dataset which comprises 450 diagnostic H&E WSIs of squamous cell carcinoma (SCC) cases from different sites of the head and neck (H&N). However, many of these slides suffer from preparation and scanning artefacts. After excluding cases with slides of poor quality, our final TCGA-HN cohort consisted of 342 cases with one WSI per case (see supplementary material, Tables S6). The H&E WSIs from these cases had mostly been scanned at 40×, with some scanned at 20×. The SKM cohort (C2) consists of 100 OSCC cases collected from the Shaukat Khanum Memorial Hospital and Research Centre (SKMCH&RC) Lahore, Pakistan [19]. The PredicTR1 cohort (C3) contains 95 OPSCC cases collected from six different centres across the UK [27]. The representative H&E tissue sections for both the SKM and the PredicTR1 cohort were all scanned at 40× magnification.

Ethics statement

Ethical approval for SKM [19] was obtained from the institutional review board (Ref. No. 17-02-17-10) at SKMCH&RC and the National Bioethics Committee (No. 4-87/17/NBC-234-Exempt/NBC/2592), Pakistan. Cases for the PredicTR1 cohort [27] were collected from six different centres after receiving approval from Research Ethics Committee (REC 11/NQ/0452) and Northern Ireland Biobank (NIB 11/001).

Survival information

DSS information was available and was obtained for most of the selected cases in all three cohorts. The DSS time was calculated from the date of diagnosis to the date of death or the date of the last follow-up in the case of censored data. DFS information was only available for the C2 and C3 cohorts. The DFS was censored at the date of first recurrence or death, whichever occurred first, or the date of the last contact for patients alive without recurrent disease. A detailed description of the patients' characteristics and available clinical and pathological parameters can be found in Supplementary materials and methods and supplementary material, Tables S1–S3.

Quantification of tumour-associated stroma infiltrating lymphocytes

We have developed an objective and automated score for the quantification of tumour-associated stroma infiltrating lymphocytes, namely the TASIL-score, which quantifies lymphocytes in the vicinity of TAS using the spatial co-occurrence statistics of both TAS and lymphocytes in a WSI. The TASIL-score is computed in two

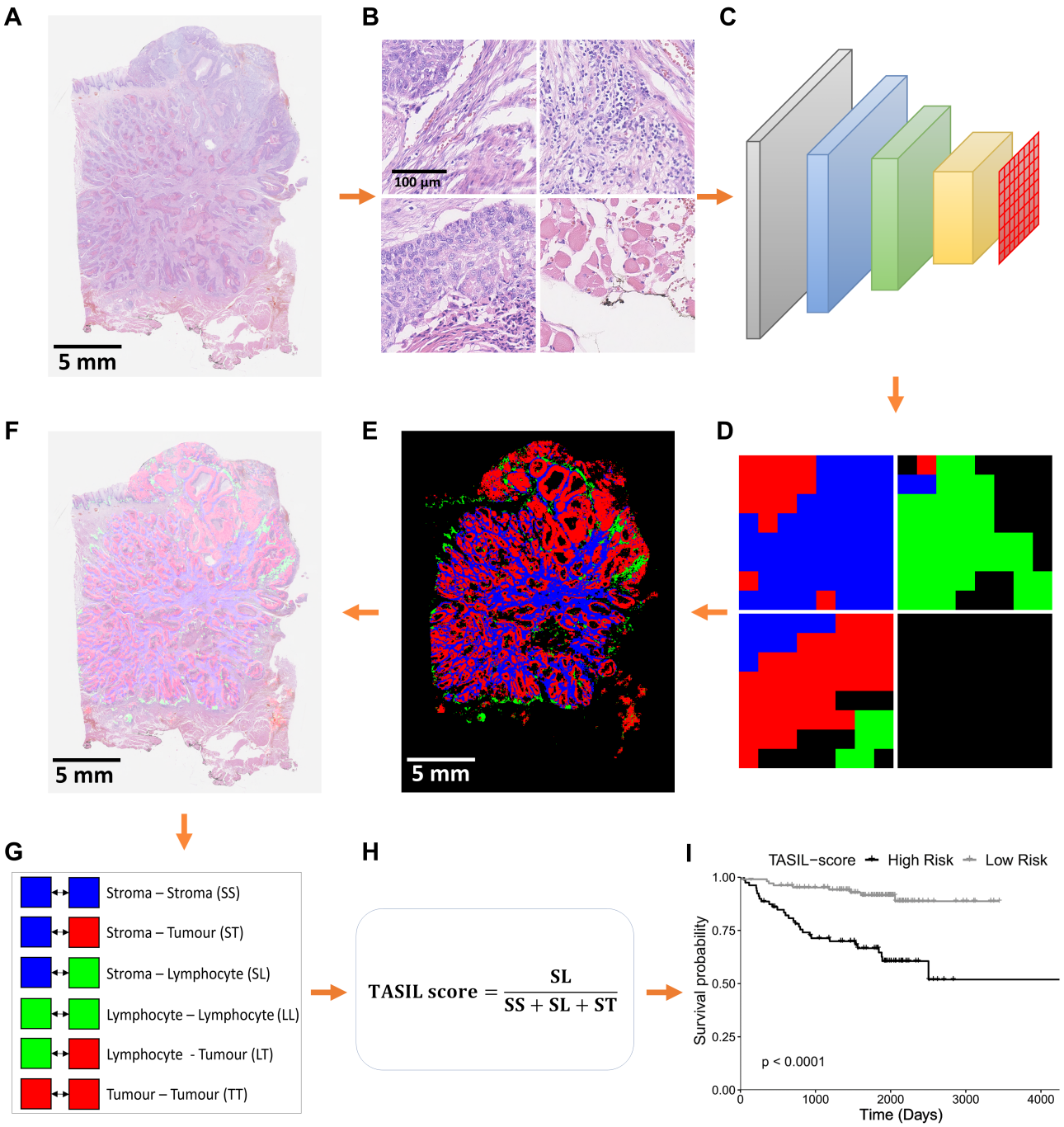

Figure 1. Flow diagram of the proposed method. (A) A whole slide image (WSI) (H&E-stained). (B) Image patches extracted from the WSI. (C) A convolutional neural network-based patch segmentation method. (D) Patch-based segmentation results, where red, green, blue, and black colours represent tumour, lymphocyte, tumour-associated stroma, and non-regions of interest, respectively. (E) Segmentation map of the WSI. (F) Overlay of the segmentation map over the WSI. (G) Analysis of the spatial co-occurrence of tumour, lymphocyte, and tumour-associated stroma regions. (H) Calculation of the proposed TASIL-score based on the number of each pair of co-occurrences. (I) TASIL-score-based patient stratification into low- and high-risk groups.

steps: segmentation of WSIs into clinically significant tissue types and calculation of the spatial co-occurrence statistics.

In the first step, we divided a given WSI into small sub-images (patches) and employed a deep learning-based patch classification method to segment the WSI into four different types of regions: tumour, TAS, lymphocytes, and all other tissue regions as non-regions of interest (non-ROIs). In the second step, spatial co-occurrence statistics were calculated using the co-occurrence analysis of different types of patches in a WSI. A patch adjacent to another patch in any direction was considered as an instance of co-occurrence. First, six different patch co-occurrence patterns were defined based on three clinically significant tissue types, as shown in supplementary

material, Figure S1. Then the TASIL-score was calculated as follows:

$$\text{TASIL} - \text{score} = \frac{SL}{SS + SL + ST}$$

where SL represents the number of times tumour-associated stroma and lymphocyte patches co-occur in a WSI. Similarly, SS and ST denote the number of co-occurrences of TAS patches with other TAS patches and tumour patches, respectively. The TASIL-score ranges from 0 to 1, where 0 represents no infiltration and 1 represents a high degree of lymphocytic infiltration in the tumour-associated stroma.

Statistical methods and data analysis

Survival analysis was performed with DSS and DFS data. The Kaplan–Meier estimator was used, and the log-rank test was performed to stratify patients into low- and high-risk groups (log-rank test-based P values were calculated to assess the significance of the various features including the proposed TASIL-score). The Cox proportional hazard regression model was used for univariate and multivariate analyses and 95% confidence intervals were computed. The Spearman rank-order correlation coefficient was used for correlation analyses between TASIL-score and molecular estimates, and between TASIL-score and the manual TIL score assigned by a pathologist. The concordance statistics were used to compare the different automated quantification scores for DSS analysis.

Supplementary materials and methods presents detailed descriptions of patients' characteristics, data annotations, and the artificial intelligence (AI) method.

Results

Automated segmentation of whole slide images for quantification of the TASIL-score

A deep learning-based segmentation algorithm was trained and evaluated on WSIs from the C1 and C2 cohorts, where ten WSIs were used for training and two for validation from each cohort. More than 179K patches (141K for training and 38K for validation) were annotated by an expert pathologist (SAK). The segmentation method achieved an average accuracy of 0.85 and macro F1-score of 0.83. Quantitative and visual results for WSI segmentation are presented in supplementary material, Table S5 and Figure S2. Performance of the segmentation method was further evaluated by calculating the Spearman correlation between the percentage of predicted lymphocyte patches (L-Percentage) and the pathologists' manually assigned TIL score on the C3 cohort. The pathologists scored TILs on an H&E slide under 2.5× magnification as high (diffuse, lymphocytes present in more than 80% of tumour/tumour-associated stroma), low (weak/absent, lymphocytes present in less than 20% of tumour/tumour-associated stroma), or moderate (patchy, present in 20–80% tumour/tumour-associated stroma) [28]. A Spearman correlation score of 0.71 with a highly significant P value of 5.10×10^{-16} was observed between the two scores. The distribution of L-Percentage across the three groups of the pathologists' TIL score showed a clear separation (see supplementary material, Figure S3).

Higher TASIL-score is associated with better disease-specific survival of HNSCC patients

We investigated the prognostic significance of the deep learning-based TASIL-score for DSS of HNSCC

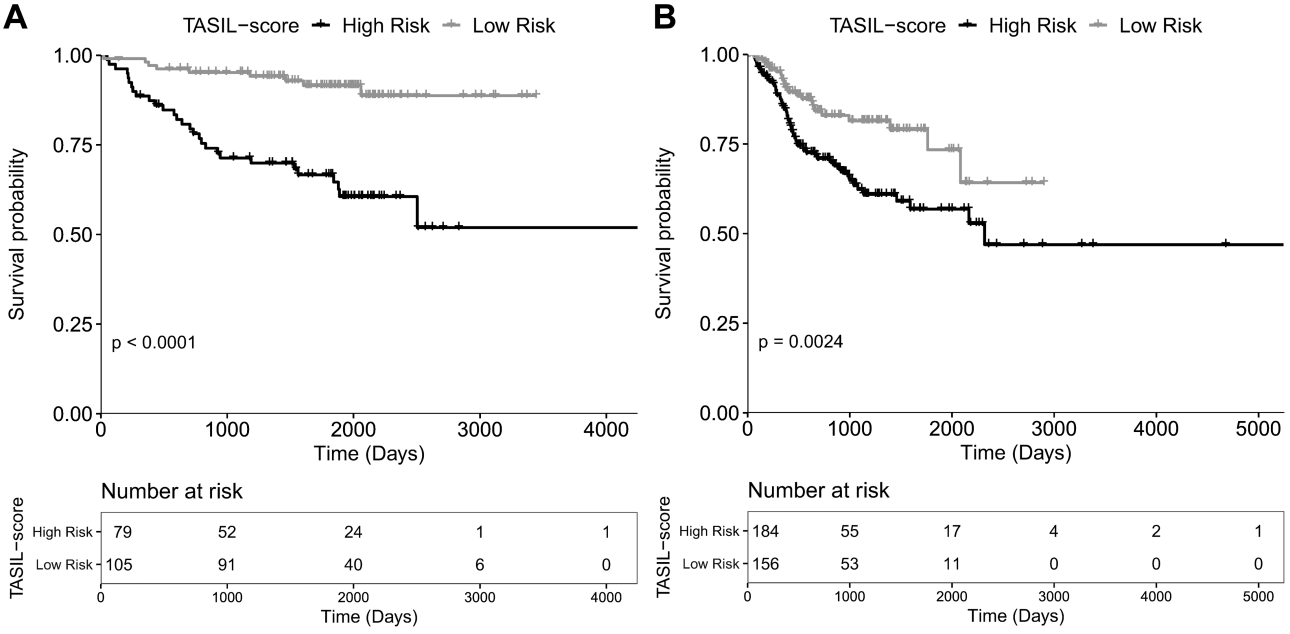

Figure 2. Kaplan–Meier curves for low- and high-risk patients in (A) the C2 and C3 cohorts and (B) the C1 cohort when used as validation cohorts for disease-specific survival (DSS).

patients. First, we considered C1 as a discovery cohort and both C2 and C3 as a joint validation cohort. Patients in the validation cohort were divided into two groups based on the TASIL-score using an optimal threshold value obtained from the analysis of the discovery cohort. We found that the patient group with the higher TASIL-score showed a significantly better DSS [$p = 0.000003$, hazard ratio (HR) = 0.20, 95% confidence interval (CI) 0.10–0.43] on the validation cohort. The TASIL-score was again statistically significant for patient stratification for DSS ($p = 0.00239$, HR = 0.49, 95% CI 0.30–0.78) when the discovery and validation cohorts were swapped. The Kaplan–Meier curves along with the corresponding log-rank test-based *P* values are presented in Figure 2. These curves show a clear separation between low- and high-risk patient groups when stratified using the TASIL-score.

TASIL-score is a prognostic indicator for the SCC of both oral and oropharynx sites

The C2 and C3 cohorts were curated from only the oral cavity and oropharynx, respectively. Therefore, we investigated the prognostic significance of TASIL-score for patients with SCC of a specific site. First, the OSCC cohort (C2) was considered as a discovery cohort and the OPSCC cohort (C3) was considered as a validation cohort. Supporting our aforementioned findings, the TASIL-score remained prognostically significant ($p = 0.000159$, HR = 0.20, 95% CI 0.08–0.49) for

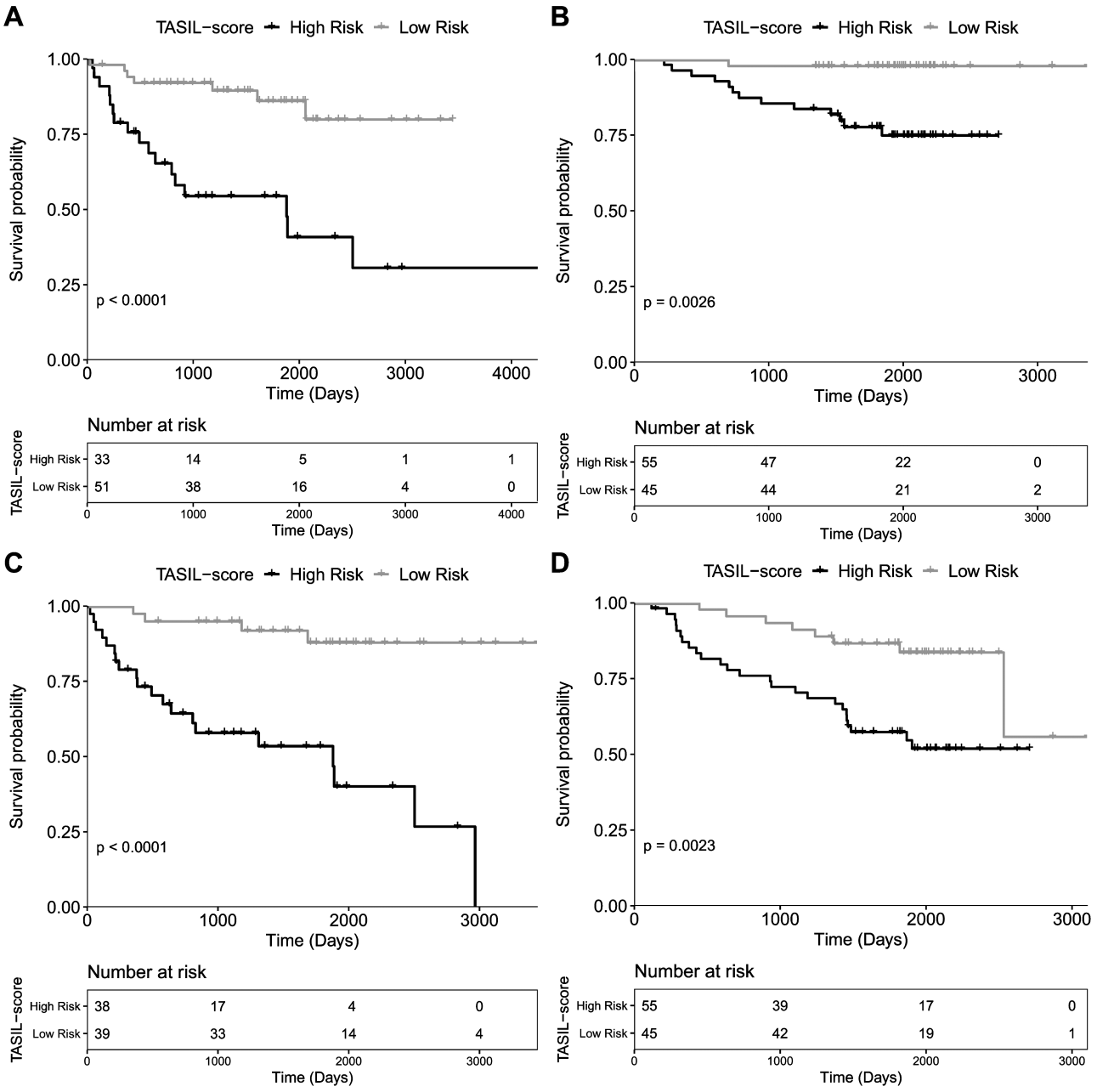

Figure 3. Kaplan–Meier curves for low- and high-risk patients in the oropharyngeal (A, C) and oral (B, D) cohorts when used as validation cohorts for disease-specific (A, B) and disease-free (C, D) survival.

Table 1. Multivariate analysis of TASIL-score in the presence of available clinical and pathological variables of the SKM cohort for both disease-specific and disease-free survival.

	Disease-specific survival				Disease-free survival			
		95% CI				95% CI		
Variable	HR	Lower	Upper	*P* value	HR	Lower	Upper	*P* value
Age	1.015	0.969	1.060	0.526	1.010	0.980	1.040	0.579
Sex	2.057	0.502	8.430	0.316	1.540	0.650	3.660	0.323
Smoking tobacco	1.905	0.505	7.190	0.342	1.160	0.510	2.690	0.72
Smokeless tobacco	0.671	0.156	2.900	0.593	1.990	0.770	5.130	0.155
Tumour grade	0.689	0.312	1.520	0.355	1.090	0.660	1.790	0.73
Invasion pattern	1.121	0.653	1.920	0.678	1.350	0.930	1.950	0.115
TNM stage	2.015	1.021	3.970	**0.043**	1.090	0.780	1.520	0.627
TASIL	0.099	0.013	0.760	**0.027**	0.290	0.120	0.670	**0.004**

Values in bold indicate statistically significant differences.

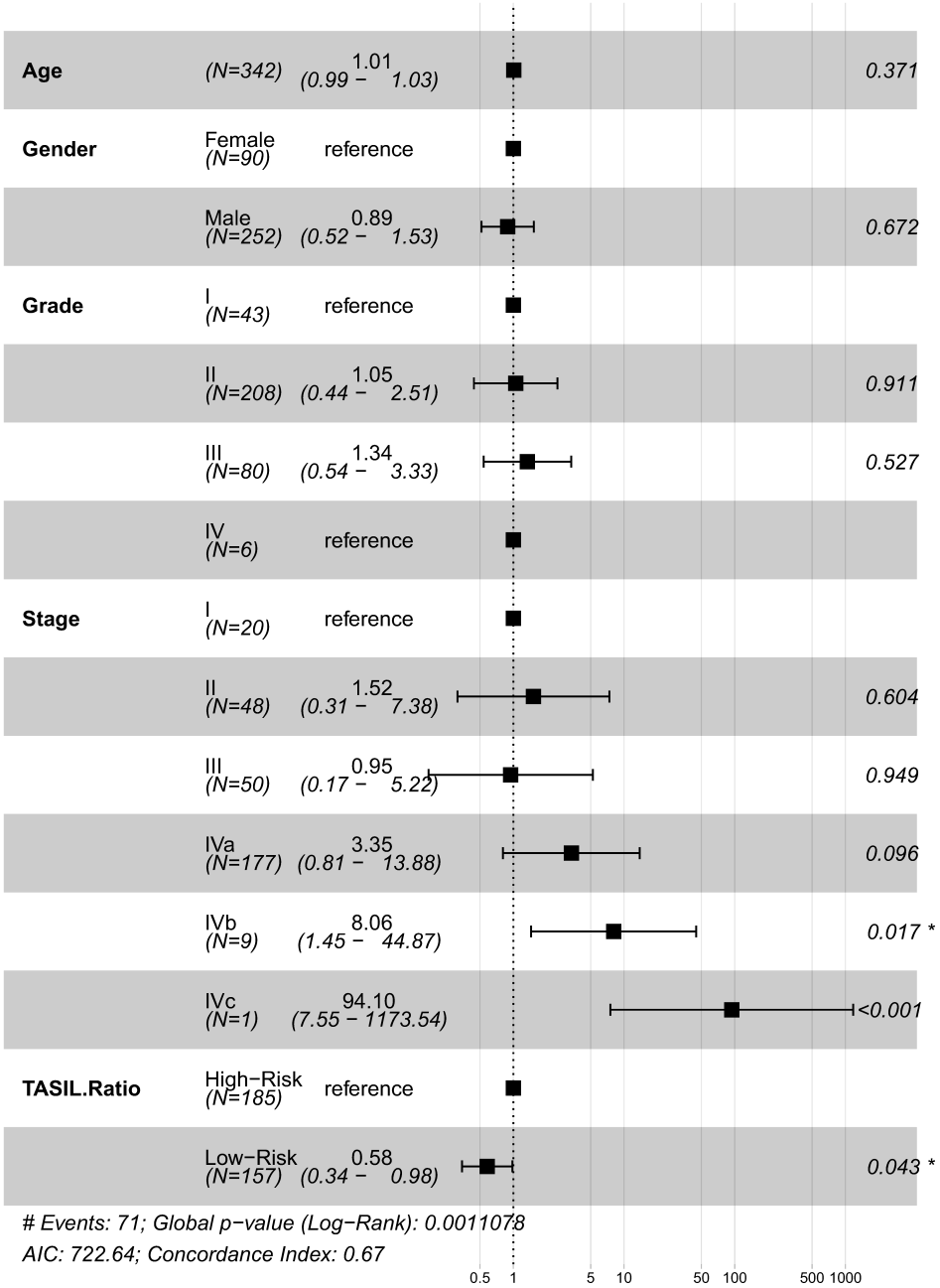

Figure 4. Multivariate analysis of TASIL-score in the presence of available clinical and pathological variables of the C1 cohort. Dotted lines represent hazard ratios.

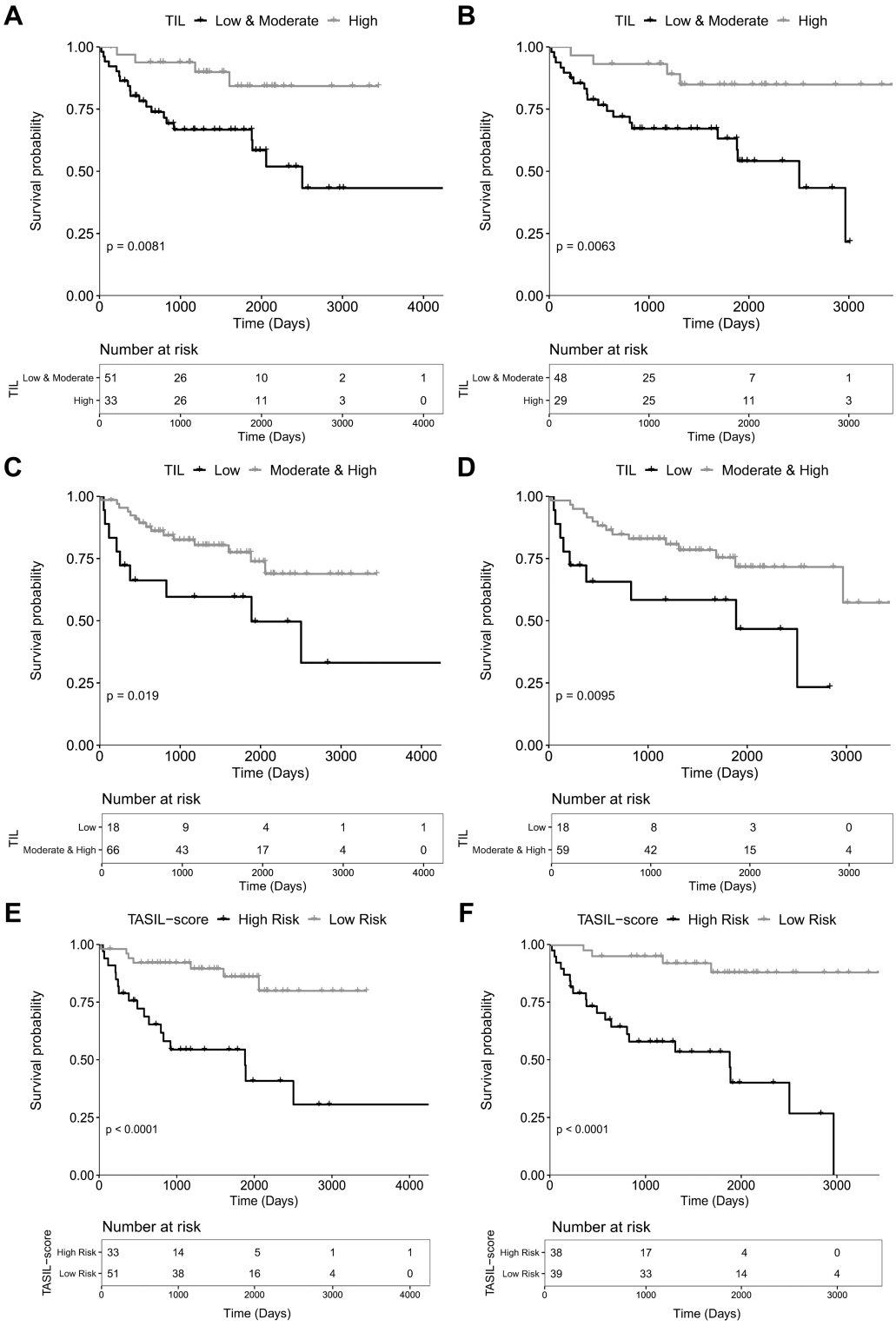

Figure 5. Comparison of the manual pathologist TIL score (A, B – Low & Moderate versus High; C, D –Low versus Moderate & High) and the proposed TASIL-score (E, F) in a univariate setting for disease-specific (A, C, E) and disease-free survival (B, D, F) of the C3 cohort.

OPSCC patient stratification into low- and high-risk groups for DSS. Second, we swapped the two cohorts, considering the OPSCC cohort (C3) as the discovery cohort and the OSCC cohort (C2) as the validation cohort. We found that the TASIL-score-based OSCC patient stratification again proved prognostically significant ($p = 0.000935$, HR = 0.08, 95% CI 0.01–0.65). When the experiments were repeated to evaluate the prognostic significance of the TASIL-score for DFS, the same pattern, where the TASIL-score stratified patients into prognostically significant low- and high-risk groups, was followed. Patient stratification into low- and high-risk groups is presented in Figure 3 through Kaplan–Meier curves for all four experiments.

Table 2. Concordance index of different tumour, tumour-associated stroma, and lymphocyte quantification methods, including the proposed TASIL-score.

Survival type	Cohort	TS-Ratio	IC-co-localization	TILAb	TASIL
Disease-specific survival	C1	0.4833	**0.6180**	0.6062	0.6146
Disease-specific survival	C2	0.5931	0.6028	0.6219	**0.6992**
Disease-specific survival	C3	0.5965	0.6228	0.5926	**0.6943**
Disease-specific survival	C2 and C3	0.5884	0.6701	0.6587	**0.7248**
Disease-free survival	C2	0.5591	0.6589	**0.6815**	0.6546
Disease-free survival	C3	0.6274	0.7113	0.7075	**0.7698**

Bold values indicate the highest concordance-index score in each row.

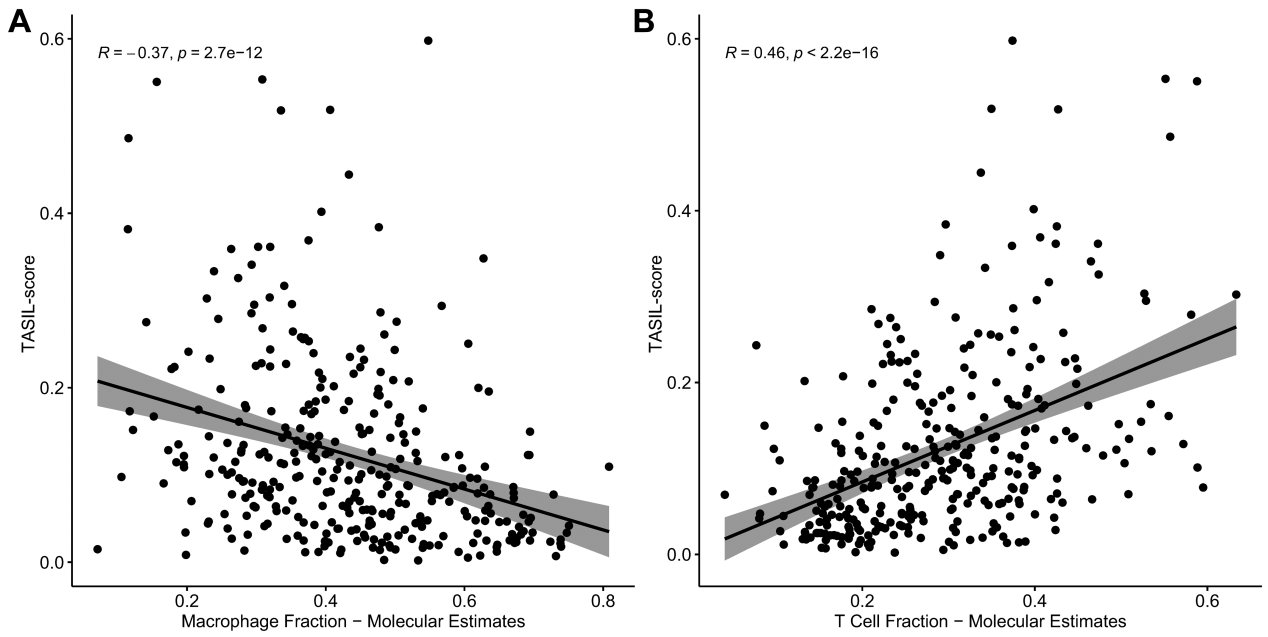

Figure 6. Spearman correlation between TASIL-score and molecular estimates of (A) macrophage and (B) T-cell fractions.

TASIL-score is independent of clinical and pathological variables

We investigated the prognostic significance of the TASIL-score through multivariate analysis in the presence of clinical and pathological variables. The C2 cohort was considered for multivariate analysis as it has more clinical and pathological parameters, along with both DSS and DFS information (supplementary material, Table S2), compared with the other two cohorts. For DSS, both TASIL-score ($p = 0.027$, HR = 0.10, 95% CI 0.01–0.76) and pathological stage ($p = 0.043$, HR = 2.02, 95% CI 1.02–3.97) were found to be independent prognostic variables against all other variables (Table 1). However, for DFS, TASIL-score was the only independent variable of statistical significance ($p = 0.004$, HR = 0.29, 95% CI 0.12–0.67) against age, sex, smoke and smokeless tobacco status, tumour grade, patterns of invasion, and pathological stage. We also investigated the prognostic significance of TASIL-score for DSS in a multivariate setting on the C1 cohort using the available clinical and pathological variables. The TASIL-score remained prognostic ($p = 0.043$, HR = 0.58, 95% CI 0.34–0.98) in the presence of other clinicopathological variables: age, gender, grade, and pathological stage. Although stages IVb and IVc appeared to be statistically significant, the total number of patients in stages IVb and IVC was 9 and 1, respectively, which is quite small compared with the total number of patients (Figure 4).

TASIL-score shows a better separation between low- and high-risk patients compared with the manual TIL score

The pathologist score for tumour/stroma infiltrating lymphocytes is usually a categorical score with low, moderate, and high infiltration categories. Only low and high categories showed prognostic significance in the C3 cohort for DSS and DFS, as shown in supplementary material, Table S3. We merged the moderate category in two different ways to split the patients into low- and high-risk groups (see Figure 5). The Kaplan–Meier curves in Figure 5 show that the TASIL-score stratifies patients into low- and high-risk groups with better prognostic significance compared with the manual TIL score in both DSS and DFS analyses.

Table 3. Spearman correlation between TASIL-score and molecular estimates of immune subtypes.

Cell type	rho	*P* value	Cell subtypes	rho	*P* value
Mast cells	−0.18	**8.22×10^{-4}**	Activated	−0.18	**1.19×10^{-3}**
			Resting	0.12	**2.84×10^{-2}**
Monocytes	0.18	**7.70×10^{-4}**	Monocytes	0.18	**7.70×10^{-4}**
Macrophages	−0.37	**2.67×10^{-12}**	M0	−0.39	**2.08×10^{-13}**
			M1	0.32	**2.62×10^{-9}**
			M2	−0.24	**1.05×10^{-5}**
T cells	0.46	**1.32×10^{-18}**	CD4 memory activated	0.32	**2.04×10^{-9}**
			CD4 memory resting	−0.09	**9.66×10^{-2}**
			CD4 naive	−0.20	**3.11×10^{-4}**
			CD8	0.45	**9.28×10^{-18}**
			Follicular helper	0.26	**2.37×10^{-6}**
			Gamma-delta	0.07	**1.94×10^{-1}**
			Regulatory	0.26	**1.16×10^{-6}**
B cells	0.22	**3.97×10^{-5}**	Memory	−0.02	**7.37×10^{-1}**
			Naive	0.22	**4.32×10^{-5}**
			Plasma	0.11	**4.77×10^{-2}**

Values in bold indicate statistically significant differences.

Comparison with existing digital scores

In recent years, researchers have proposed several automated quantification methods for co-localisation in different types of cancers to develop a digital prognostic biomarker [19,22,29,30]. The use of computerised methods addresses the issue of subjectivity and produces objective and reproducible quantification scores. Geessink *et al* [30] presented tumour to stroma ratio (TS-Ratio) for rectal adenocarcinoma and showed that a higher TS-Ratio had prognostic association with poor patient survival. Maley *et al* [22] used an ecological measure for quantification of co-localisation between immune cells and cancer cells (IC-Co-localisation) in breast cancer. They found that higher co-localisation of immune and tumour cells was associated with better patient survival.

Similarly, Shaban *et al* [19] proposed a tumour-infiltrating lymphocytes abundance (TILAb) score for OSCC which showed prognostic significance for DFS in both univariate and multivariate analysis. We used Harrell's concordance index (C-Index) to compare the predictive ability of the TASIL-score with the existing automated quantification methods in six different experiments. Table 2 presents C-Index scores on all cohorts when used as validation. The proposed TASIL-score achieved the best C-Index scores in four experiments and comparable C-Index scores in the remaining two experiments for DSS and DFS.

TASIL-score is correlated with molecular estimates of $CD8^+$ T cells

We further investigated the correlation of the proposed TASIL-score with molecular estimates of immune cell fractions in the TCGA-HN cohort (C1). Thorsson *et al* [31] have estimated the fraction of 22 immune cell types in the histology slides of patients in the TCGA cohort using gene expression data through CIBERSORT. We used those estimates for the correlation analysis with our TASIL-score. The immune subtypes were grouped based on nine different immune cell types: dendritic, mast, neutrophils, eosinophils, monocytes, macrophages, natural killer cells, T cells, and B cells. Our main finding was that the TASIL-score shows a moderate but highly significant positive correlation with T-cell estimates and a negative correlation with macrophage estimates (Figure 6).

Moreover, $CD8^+$ T-cell fraction shows the highest positive correlation among all immune subtypes (Table 3), which may indicate that the lymphocytes in the vicinity of the TAS are mainly $CD8^+$ T cells [32]. An explanation for the value of the correlation between TASIL-score and cell fractions not being very high is that TASIL-score and the molecular estimates are computed on formalin-fixed, paraffin-embedded (FFPE) and fresh frozen tissue sections, respectively. Both tissue sections belong to tissue blocks from the same patient, but their exact spatial relation is unknown. However, a good correlation of TASIL-score with $CD8^+$ T cells indirectly validates the significance of the proposed score.

Discussion

We have proposed a deep learning-based objective measure, namely the TASIL-score, for quantification of tumour-associated stroma infiltrating lymphocytes in digitised images of HNSCC tissue slides. We found that a higher value of the TASIL-score was associated with better DSS of HNSCC patients (Figure 2) and with both DSS and DFS of OSCC and OPSCC patients (Figure 3). The TASIL-score was independent of clinicopathological parameters for DSS and DFS of OSCC patients (Table 1). It also showed a better separation between low- and high-risk OPSCC patients compared with the manual TIL score assessed by an expert pathologist (Figure 5). We compared the TASIL-score with the existing automated quantification methods for stroma

and lymphocytic quantification with respect to the tumour. The TASIL-score achieved a high concordance score compared with its counterparts (Table 2) and also showed a moderate but highly significant correlation with molecular estimates of $CD8^+$ T cells (Table 3).

Most of the existing automated quantification methods were developed for the quantification of lymphocytes or stroma in relation to the tumour, such as stroma to tumour ratio in breast and ovarian cancer [29,30], lymphocyte and tumour co-localisation in breast cancer [22], and abundance of tumour-infiltrating lymphocytes in OSCC [19]. However, automated quantification of stromal TILs has not been fully explored yet. To the best of our knowledge, the proposed TASIL-score is the first automated quantitative score of lymphocytic infiltration in the TAS of HNSCC.

The role of TASIL has been investigated in several clinical studies of different cancers [10,24,25]. Salgado *et al* [10] found that stromal TILs are a superior and more reproducible prognostic parameter than intratumoral TILs in breast cancer. Xu *et al* [25] also reported that stromal TILs were of clinical relevance as a high stromal TIL score was associated with better patient prognosis for HNSCC. Furthermore, stromal TILs have been shown to be an independent risk factor for DSS and DFS of HNSCC patients. Our automated TASIL-score shows a similar prognostic significance pattern (see Figures 2 and 3).

In clinical practice, both H&E staining and immunohistochemistry (IHC) staining are used for manual TIL scoring. The TASIL-score-based results are in agreement with previous findings based on H&E- and IHC-based TIL quantification [24,33,34]. In Figure 5, we have shown that both the manual H&E-based pathologist score and the TASIL-score carry prognostic significance for DSS and DFS of OPSCC. However, the TASIL-score shows better separation between Kaplan–Meier curves of low- and high-risk patients of OPSCC. We also visually analysed two cases with low and high TASIL-score from the group of TIL-High and TIL-Low cases, respectively (supplementary material, Figures S5 and S6). Visual analysis of supplementary material, Figure S5 reveals that the difference between the TASIL score and the pathologist's manual TIL score is largely due to the difference in their definitions. The proposed digital TASIL-score quantifies lymphocyte infiltration in tumour-associated stroma (TAS), whereas the manual TIL score attempts to quantify lymphocyte infiltration in both tumour and TAS, where applicable. de Ruiter *et al* [33] conducted a meta-analysis of IHC-based studies to investigate the prognostic value of T cells in HNSCC. They found a favourable prognostic role of $CD3^+$ and $CD8^+$ T-cell infiltration in HNSCC patients. Balermpas *et al* [24] found that high CD8 expression in tumour stroma is a prognosticator for HNSCC. The proposed TASIL-score also shows a positive and highly significant correlation with genomic estimates of $CD8^+$ T cells in HNSCC patients in the TCGA-HNSCC cohort.

One of the limitations of this work is the marked difference in low- and high-risk groups across different cohorts. The separation between low- and high-risk groups in the C2 and C3 cohorts is more sustained compared with the C1 cohort. The slight difference in performance on C1 may be attributed to stain variation and scanner differences, known as the domain shift problem in the area of computational pathology [35–37]. While we have shown the TASIL-score to be prognostically significant across all three cohorts (including C1), the score must be evaluated for robustness to stain variation and scanner differences before it can be deployed in clinical practice. Moreover, image quality in C1 is low compared with C2 and C3, which may result in poor segmentation of tumour, tumour-associated stroma, and lymphocytic regions (see supplementary material, Figures S7–S9). Another limitation of this work is the moderate correlation between the TASIL-score and molecular estimates of $CD8^+$ T cells. Further investigation is required to improve the correlation by including other co-occurrence patterns in the TASIL-score, e.g. lymphocyte to lymphocyte co-occurrence. A higher correlation between tailored TASIL-score and the $CD8^+$ T cells will also justify the TASIL-score's better survival stratification than the manual TIL score. However, the investigation of tailored TASIL-score with better correlation with $CD8^+$ T cells is beyond the scope of this study and a good future direction.

T-lymphocyte infiltration in the stroma and tumour indicates an effective immune challenge to the tumour and is related to better outcome and treatment response. The proposed TASIL-score-based findings are aligned with the clinical knowledge, with the added advantages of objectivity, reproducibility, and strong prognostic value. Therefore, the automated, objective, and quantitative TASIL-score has the potential to provide valuable insights into tumour behaviour and prognosis in an efficient and consistent manner. Although we validated our method on three different cohorts ($n = 537$ cases in total), a comprehensive evaluation on large multicentric cohorts is required before the proposed digital score can be adopted in clinical practice.

Acknowledgements

This study was supported by Medical Research Council UK grant number MR/P015476/1.

Author contributions statement

NR, SAK and HM conceived the experiments. MS, SAR, SAK and NR carried out experiments and analysed data. SAK, MR and HM contributed pathology expert knowledge to experimental design and data interpretation. MH, AJ, SM, AL, NB, JB and PN provided essential resources. All the authors were involved in interpreting the results, writing the paper and had final approval of the submitted and published versions.

References

1. Bray F, Ferlay J, Soerjomataram I, *et al*. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018; **68:** 394–424.
2. Jemal A, Siegel R, Ward E, *et al*. Cancer statistics, 2007. *CA Cancer J Clin* 2007; **57:** 43–66.
3. Peter B & Bernard L World Cancer Report 2008. [Accessed 26 November 2020]. Available from: https://publications.iarc.fr/Non-Series-Publications/World-Cancer-Reports/World-Cancer-Report-2008.
4. Shaw R, Beasley N. Aetiology and risk factors for head and neck cancer: United Kingdom National Multidisciplinary Guidelines. *J Laryngol Otol* 2016; **130:** S9–S12.
5. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Alcohol consumption and ethyl carbamate. *IARC Monogr Eval Carcinog Risks Hum* 2010; **96:** 3–1383.
6. Head and Neck Cancers Incidence Statistics – Cancer Research UK. [Accessed 5 February 2021]. Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/head-and-neck-cancers/incidence##heading-Four.
7. Braakhuis BJ, Brakenhoff RH, Leemans CR. Second field tumors: a new opportunity for cancer prevention? *Oncologist* 2005; **10:** 493–500.
8. Peled M, Onn A, Herbst RS. Tumor-infiltrating lymphocytes – location for prognostic evaluation. *Clin Cancer Res* 2019; **25:** 1449–1451.
9. Zhou C, Wu Y, Jiang L, *et al*. Density and location of $CD3^+$ and $CD8^+$ tumor-infiltrating lymphocytes correlate with prognosis of oral squamous cell carcinoma. *J Oral Pathol Med* 2018; **47:** 359–367.
10. Salgado R, Denkert C, Demaria S, *et al*. The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. *Ann Oncol* 2015; **26:** 259–271.
11. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, *et al*. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 2017; **318:** 2199–2210.
12. Campanella G, Hanna MG, Geneslaw L, *et al*. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med* 2019; **25:** 1301–1309.
13. Lu MY, Williamson DFK, Chen TY, *et al*. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat Biomed Eng* 2021; **5:** 555–570.
14. Qaiser T, Tsang YW, Taniyama D, *et al*. Fast and accurate tumor segmentation of histology images using persistent homology and deep convolutional features. *Med Image Anal* 2019; **55:** 1–14.
15. Lin H, Chen H, Graham S, *et al*. Fast ScanNet: fast and dense analysis of multi-gigapixel whole-slide images for cancer metastasis detection. *IEEE Trans Med Imaging* 2019; **38:** 1948–1958.
16. Pantanowitz L, Quiroga-Garza GM, Bien L, *et al*. An artificial intelligence algorithm for prostate cancer diagnosis in whole slide images of core needle biopsies: a blinded clinical validation and deployment study. *Lancet Digit Health* 2020; **2:** e407–e416.
17. Shaban M, Awan R, Fraz MM, *et al*. Context-aware convolutional neural network for grading of colorectal cancer histology images. *IEEE Trans Med Imaging* 2020; **39:** 2395–2405.
18. Lu C, Koyuncu C, Corredor G, *et al*. Feature-driven local cell graph (FLocK): new computational pathology-based descriptors for prognosis of lung cancer and HPV status of oropharyngeal cancers. *Med Image Anal* 2021; **68:** 101903.
19. Shaban M, Khurram SA, Fraz MM, *et al*. A novel digital score for abundance of tumour infiltrating lymphocytes predicts disease free survival in oral squamous cell carcinoma. *Sci Rep* 2019; **9:** 13341.
20. Bankhead P, Fernández JA, McArt DG, *et al*. Integrated tumor identification and automated scoring minimizes pathologist involvement and provides new insights to key biomarkers in breast cancer. *Lab Invest* 2018; **98:** 15–26.
21. Saltz J, Gupta R, Hou L, *et al*. Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. *Cell Rep* 2018; **23:** 181–193.e7.
22. Maley CC, Koelble K, Natrajan R, *et al*. An ecological measure of immune-cancer colocalization as a prognostic factor for breast cancer. *Breast Cancer Res* 2015; **17:** 131.
23. Nawaz S, Heindl A, Koelble K, *et al*. Beyond immune density: critical role of spatial heterogeneity in estrogen receptor-negative breast cancer. *Mod Pathol* 2015; **28:** 1621.
24. Balermpas P, Michel Y, Wagenblast J, *et al*. Tumour-infiltrating lymphocytes predict response to definitive chemoradiotherapy in head and neck cancer. *Br J Cancer* 2014; **110:** 501–509.
25. Xu Q, Wang C, Yuan X, *et al*. Prognostic value of tumor-infiltrating lymphocytes for patients with head and neck squamous cell carcinoma. *Transl Oncol* 2017; **10:** 10–16.
26. The Cancer Genome Atlas Home Page NIH: National Cancer Institute. U.S. Department of Health and Human Services. [Accessed 5 February 2021]. Available from: https://cancergenome.nih.gov/.
27. Schache AG, Powell NG, Cuschieri KS, *et al*. HPV-related oropharynx cancer in the United Kingdom: an evolution in the understanding of disease etiology. *Cancer Res* 2016; **76:** 6598–6606.
28. Ward MJ, Thirdborough SM, Mellows T, *et al*. Tumour-infiltrating lymphocytes predict for outcome in HPV-positive oropharyngeal cancer. *Br J Cancer* 2014; **110:** 489–500.
29. Kemi N, Eskuri M, Herva A, *et al*. Tumour-stroma ratio and prognosis in gastric adenocarcinoma. *Br J Cancer* 2018; **119:** 435–439.
30. Geessink OGF, Baidoshvili A, Klaase JM, *et al*. Computer aided quantification of intratumoral stroma yields an independent prognosticator in rectal cancer. *Cell Oncol (Dordr)* 2019; **42:** 331–341.
31. Thorsson V, Gibbs DL, Brown SD, *et al*. The immune landscape of cancer. *Immunity* 2018; **48:** 812–830.e14.
32. AbdulJabbar K, Raza SEA, Rosenthal R, *et al*. Geospatial immune variability illuminates differential evolution of lung adenocarcinoma. *Nat Med* 2020; **26:** 1054–1062.
33. de Ruiter EJ, Ooft ML, Devriese LA, *et al*. The prognostic role of tumor infiltrating T-lymphocytes in squamous cell carcinoma of the head and neck: a systematic review and meta-analysis. *Onco Targets Ther* 2017; **6:** e1356148.
34. Fang J, Li X, Ma D, *et al*. Prognostic significance of tumor infiltrating immune cells in oral squamous cell carcinoma. *BMC Cancer* 2017; **17:** 375.
35. Lafarge MW, Pluim JPW, Eppenhof KAJ, *et al*. Domain-adversarial neural networks to address the appearance variability of histopathology images. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Lecture Notes in Computer Science , Cardoso MJ, Arbel T, Carneiro G, *et al*. (eds). Springer International Publishing: Cham, 2017; 83–91.
36. Koohbanani NA, Unnikrishnan B, Khurram SA, *et al*. Self-Path: self-supervision for classification of pathology images with limited annotations. *IEEE Trans Med Imaging* 2021; **40:** 2845–2856.
37. Foote A, Asif A, Azam A, *et al*. Now you see it, now you don't: adversarial vulnerabilities in computational pathology. arXiv 2021; 2106.08153. Available from: https://arxiv.org/abs/2106.08153v2 Not peer reviewed.

SUPPLEMENTARY MATERIAL ONLINE

Supplementary materials and methods